

# Size-invariant descriptors for detecting regions of abnormal growth in cervical vertebrae

R. Joe Stanley<sup>a,\*</sup>, Sameer Antani<sup>b</sup>, Rodney Long<sup>b</sup>, George Thoma<sup>b</sup>,  
Kapil Gupta<sup>a</sup>, Mohammed Das<sup>c</sup>

<sup>a</sup> Department of Electrical and Computer Engineering, University of Missouri-Rolla, Rolla, MO, United States

<sup>b</sup> Communications Engineering Branch, National Library of Medicine, Bethesda, MD, United States

<sup>c</sup> Department of Computer Science, University of Missouri-Rolla, Rolla, MO, United States

Received 26 January 2007; received in revised form 29 August 2007; accepted 4 September 2007

---

## Abstract

Digitized spinal X-ray images exhibiting specific pathological conditions such as osteophytes can be retrieved from large databases using Content Based Image Retrieval (CBIR) techniques. For efficient image retrieval, it is important that the pathological features of interest be detected with high accuracy. In this study, new size-invariant features were investigated for the detection of anterior osteophytes, including claw and traction in cervical vertebrae. Using a K-means clustering and nearest neighbor classification approach, average correct classification rates of 85.80%, 86.04% and 84.44% were obtained for claw, traction and anterior osteophytes, respectively.

*Keywords:* CBIR; Cervical spine; Claw; Image processing; K-means clustering; Nearest neighbor classification; Osteoarthritis; Osteophyte; Traction; X-ray; NHANES

---

## 1. Introduction

Osteoarthritis, also known as degenerative joint disease, is an orthopedic anomaly affecting millions of Americans, with people over the age of 75 exhibiting increased vulnerability. The condition arises as a result of thinning of cartilage tissue covering the bone joints in the human body, thereby increasing friction during joint movement and causing a sensation of pain. The joints affected by osteoarthritis often exhibit abnormal bone growth, resulting in formation of “bone spurs”, also known as osteophytes. Radiographs provide a fast and practical approach for visualization of features such as osteophytes, disc space narrowing, and subluxation, which are of great interest to the osteoarthritis research community. Fig. 1 presents a cervical spine X-ray image example. The boxed region highlights the cervical spine vertebrae.

The Lister Hill National Center for Biomedical Communications, an R&D division of the National Library of Medicine, National Institutes of Health, has been actively conducting and

promoting research in the area of computer-assisted analysis of spine X-ray images and has developed the Web-based Medical Information Retrieval System (WebMIRS). This system provides online access to a large database of spine X-ray images and related textual data collected as a part of the National Health and Nutrition Examination Surveys (NHANES). Content Based Image Retrieval (CBIR) techniques can be used to retrieve digitized radiographs of the spine which exhibit one or more specific physiological conditions such as the presence of anterior osteophytes. The reliability of the retrieval process depends on the accuracy with which the pathology sought can be detected. This research focuses on the computer-assisted discrimination of variations of anterior osteophytes in normal cervical spine vertebrae.

Osteophytes manifest themselves as deviations from normal shape of the affected vertebra in certain specific locations. Information from other locations on the vertebra, where the shape is normal, is irrelevant and serves to degrade the efficiency of the retrieval process. Dynamic programming-aided partial shape matching techniques have been used by Xu et al. to detect anterior osteophytes [1,2].

Alternative schemes to classify anterior osteophytes have been developed, including Macnab’s classification based on

---

\* Corresponding author.

E-mail address: stanleyr@umr.edu (R.J. Stanley).

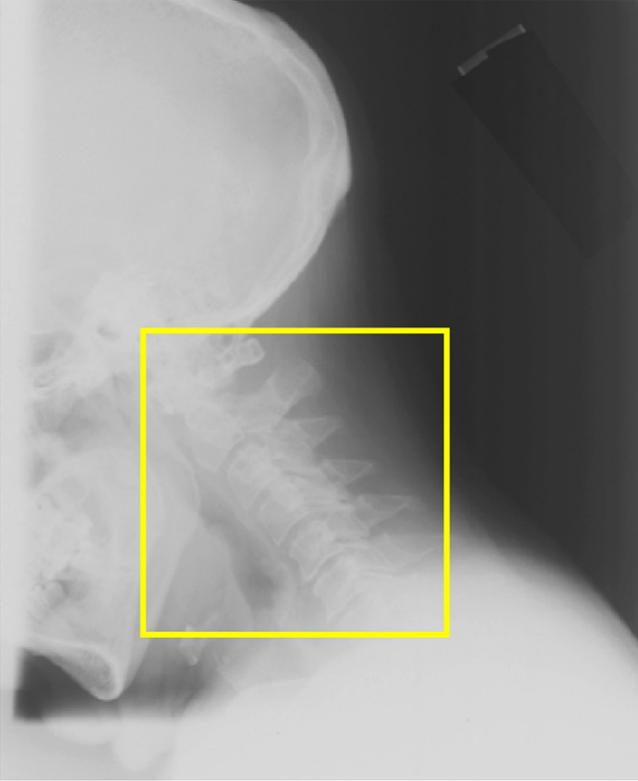


Fig. 1. Cervical spine X-ray image example from the NHANES image collection archived at the National Library of Medicine. Vertebrae are highlighted in the box region.

radiology and pathology [3,4], and a grading system defined by a medical expert to assign severity levels to the Macnab classes. Macnab’s classification defines claw and traction osteophytes. A claw osteophyte extends from the vertebral rim and curves in the direction of the adjacent disc. A claw region is typically triangular in shape and is curved at the tip of the region. A traction osteophyte tends to protrude horizontally, is usually thick, does not tend to curve at the tips and does not extend across the intervertebral disc space. The severity grading system includes three grades for osteophytes: slight, moderate and severe. If a vertebra does not exhibit claw or traction or slight, moderate or severe grades, the vertebra is considered normal.

Fig. 2 provides borders of cervical vertebrae C3–C6, as determined by an expert at the National Library of Medicine (NLM). The top of each vertebra is referred to as the superior side, and the bottom as the inferior side. The left side of the vertebra is the anterior side, and the right side is the posterior side (along the spinal column). Table 1 shows the verified claw, traction, and anterior osteophytes grades for the vertebrae examples in Fig. 2 for the inferior and superior sides of the each vertebra. Data “truthing” for claw, traction, and anterior osteophytes was performed by an expert from NLM.

The variable quality of the spine X-ray images in the NHANES data set makes it difficult to detect certain subtle pathologies and also results in low inter/intra observer repeatability. The use of relevance feedback along with partial shape matching techniques has been investigated to help refine the con-

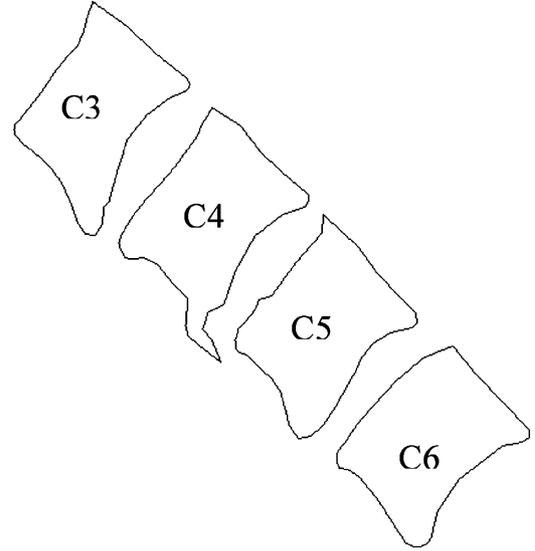


Fig. 2. Image examples of cervical vertebrae C3–C6.

tent based retrieval process by incorporating judgments made by a skilled human observer [5]. Antani et al. have investigated partial shape matching retrieval techniques. These researchers used information about vertebral boundary semantics determined from an automated localization algorithm based on nine boundary points marked by a skilled radiologist [6]. Also examined are deformable shape models that have been examined and are capable of representing globular shapes as well as subtle localized variations in features for vertebra segmentation [7].

In previous research, we investigated four size-invariant convex hull features to discriminate of anterior osteophytes in lumbar vertebrae [8]. The convex hull of a set of points is the smallest convex set that includes all of the points in the original set [9]. The shape of a normal vertebra is convex and nearly rectangular and is expected to be very similar to the convex hull constructed from its boundary points. The presence of anterior osteophytes results in deviation of the vertebra from its characteristic convex shape. The convex hull features provide a means to quantify the variation in a vertebra’s shape from a typical convex shape and to identify and quantify protrusion regions on the vertebra’s anterior side; these regions of protrusion are characteristic of anterior osteophytes. Fig. 3 provides examples of images that were used for calculation of the four convex hull-based features.

Table 1

Claw, traction and anterior osteophytes labels and grades for the inferior and superior sides of each vertebra shown in the cervical images from Fig. 1

Vertebra	Location	Claw	Traction	Anterior osteophytes grade
C3	Inferior/superior	True/false	True/false	Moderate/slight
C4	Inferior/superior	True/false	False/true	Severe/slight
C5	Inferior/superior	False/false	True/false	Slight/slight
C6	Inferior/superior	False/false	False/false	Slight/slight

The claw and traction labels are True or False for the inferior and superior sides, and the anterior osteophytes grades are slight, moderate or severe for the inferior and superior sides.

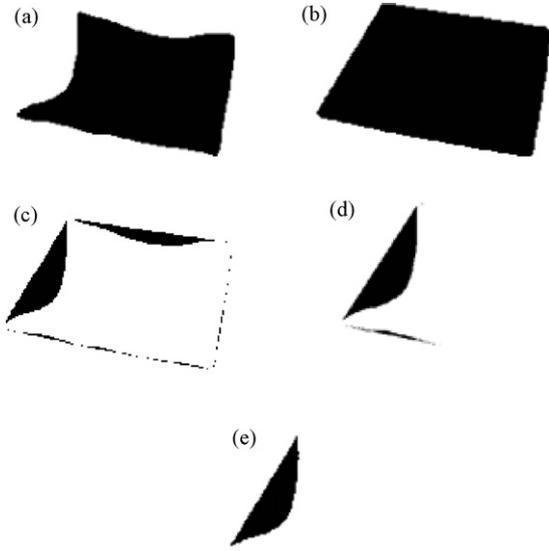


Fig. 3. Examples of images used for convex hull-based feature calculations for an abnormal vertebra: (a) filled vertebra image; (b) filled convex hull image; (c) exclusive-OR image of (a) and (b); (d) exclusive-OR image region on vertebra's anterior side; (e) connected component with largest area on anterior side.

In this study, new size-invariant features are proposed for cervical vertebrae analysis, including anterior osteophytes discrimination and the detection of claw and traction. The proposed features extend previous research [8] to detect anterior osteophytes, which utilized size-invariant-based descriptors to quantify deviations of a vertebra's shape from its typical convex shape. The extended size-invariant features quantify the vertebral deviation from convexity on the superior, inferior and anterior sides. K-means clustering and nearest neighbor methods are used for feature discrimination on a dataset of 390 cervical vertebrae. The remainder of the paper presents the: (1) algorithms for feature calculations, (2) experiments performed, (3) experimental results and discussion, and (4) conclusions.

## 2. Feature calculation algorithms for discrimination of claw, traction and anterior osteophytes in cervical spine

In this study, size-invariant features were explored to discriminate normal vertebrae and abnormal ones containing claw, traction and anterior osteophytes. The following sections describe, in detail, the algorithm for vertebral boundary determination and vertebral preprocessing prior to feature calculations. Algorithms for computing the six proposed features are then presented.

### 2.1. Vertebral boundary determination and preprocessing algorithm

The data set used in this research, supplied by the National Library of Medicine, consisted of text files for each cervical vertebra containing 36 coordinates along the vertebra's boundary marked by experienced radiologists. The first step in the preprocessing algorithm for determining the convex hull-based

features was to generate a connected boundary for the vertebra from the 36 coordinates by applying a B-spline technique [8]. Second, the connected boundary was filled to generate a vertebra for processing. Let  $D = D(x,y)$  denote the filled vertebrae with area  $A_D$  such that

$$D = \begin{cases} 1, & \text{if } (x, y) \text{ is on or inside the vertebra boundary} \\ 0, & \text{otherwise} \end{cases}$$

Third, the convex hull of  $D$  was determined using the quick convex hull algorithm [9]. Let  $H = H(x,y)$  denote the resulting filled convex hull for vertebra  $D$  with area  $A_H$  such that

$$H = \begin{cases} 1 & \text{if } (x, y) \text{ is on or inside the convex hull boundary for } D \\ 0 & \text{otherwise} \end{cases}$$

Let  $X$  denote the set of exclusive-OR points between  $D$  and  $H$  such that  $X(x,y) = D(x,y) \oplus H(x,y)$ .  $X$  is expected to contain one or more connected regions, for each concave vertebral side representative of abnormality in vertebral shape. Let  $A_X$  denote the area of the exclusive-OR region. Using 8-connectivity, let  $k$  denote the number of unique connected components within  $X$ . Fourth, the centroid  $(\bar{x}_D, \bar{y}_D)$  of the filled vertebra, and the centroids  $(\bar{x}_i, \bar{y}_i)$ ,  $i = 1, \dots, k$ , of the  $k$  connected components with areas  $A_1, \dots, A_k$  within  $X$  were calculated. Fifth, the posterior side of the vertebra within the exclusive-OR image  $X$  was eliminated, since we are only analyzing anterior, superior, and inferior side features and the posterior side information is irrelevant. The posterior side of the vertebra was identified from the image  $X$  as the set  $M = \{m_1, \dots, m_n\}$  of connected components with centroids  $(\bar{x}_j, \bar{y}_j)$  with coordinates greater than  $\bar{x}_D - R$  and less than  $\bar{y}_D - Q$ , where  $R$  and  $Q$  were small offsets from the centroid positions. Fig. 4 presents an example of bounding the posterior side of a vertebra, as shown with dashed lines. The area of the posterior side was calculated as

$$A_P = \sum_{\substack{j=1 \\ m_j \in M}}^n A_{m_j},$$

where  $A_{m_j}$  denotes the area of the  $m_j$ th connected component. Sixth, the area of the inferior side, denoted as  $A_I$ , was computed using a similar coordinate offset approach to isolate that side's connected components. Seventh, the area  $A_S$  of the superior side, was calculated using an analogous procedure. Eighth, the area  $A_T$  of the anterior side of the vertebra was similarly calculated. Ninth, the orientation of the vertebra was estimated from the line connecting corner points of the posterior side. The corner points of the posterior side correspond to the topmost and rightmost points of the vertebra. Tenth, the vertebra was rotated by the angle estimated from the posterior side to an approximately horizontal orientation. Eleventh, moment normalization [10] was performed to facilitate vertical alignment of the anterior and posterior sides of the vertebra. Twelfth, the area  $A_Z$  of the moment-normalized vertebra was computed. Thirteenth, the anterior side of the vertebra was flipped across the horizontal axis going through the centroid of the moment-normalized vertebra,

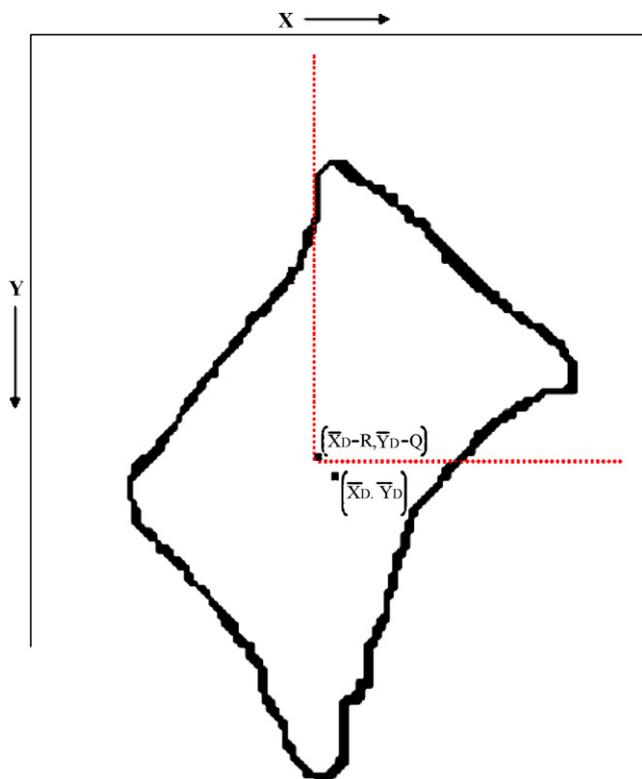


Fig. 4. Image of a cervical vertebra with its posterior side bounded by dashed lines.

and the exclusive-OR area  $A_F$  between the posterior and flipped anterior sides was computed. The areas of the concave portions of the vertebra on the inferior, superior, and anterior side, as well as the moment normalized vertebra regions, provided the basis for feature calculations.

## 2.2. Description of size-invariant features

The following vertebra descriptors were calculated: (1) the ratio of the area of the filled vertebra to the area of the filled convex hull as given by the equation  $F_1 = A_D/A_H$ , (2) the ratio of the area of the exclusive-OR image with the posterior area removed to the area of the filled convex hull as expressed by  $F_2 = (A_X - A_P)/A_H$ , (3) the ratio of the area of the inferior side of the vertebra to the area of the filled vertebra as given by  $F_3 = A_I/A_D$ , (4) the ratio of the area of the superior side of the vertebra to the area of the filled vertebra as expressed by  $F_4 = A_S/A_D$ , (5) the ratio of the area of the anterior side to the area of the filled vertebra as provided by  $F_5 = A_T/A_D$ , and (6) the ratio of the area of the exclusive-OR of the moment normalized flipped anterior side with the posterior side to the area of moment normalized vertebra as given by  $F_6 = A_F/A_Z$ . An image illustration for a vertebra is presented in Fig. 5. Fig. 5(a) shows the image of the filled C3 vertebra from Figs. 2 and 5(b) shows the image of the filled convex hull of the vertebra shown in Fig. 5(a). Fig. 5(c) shows the anterior, superior and inferior sides of the regions extracted by XOR-ing the filled vertebra image with its filled convex hull image. Fig. 5(d) shows the moment-normalized vertebra.

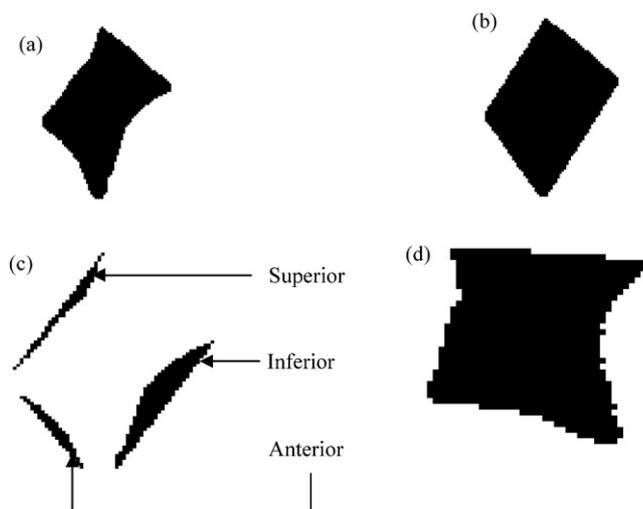


Fig. 5. Example of extraction of inferior, superior, anterior, posterior sides of vertebra from exclusive-OR of filled vertebra and associated convex hull images for C3 from Fig. 3(a): (a) filled vertebra image; (b) filled convex hull image; (c) inferior, superior and posterior sides; (d) moment normalized image.

## 3. Experiments performed

The experimental data set for evaluating the proposed features was provided by the National Library of Medicine and contained: (1) a database table listing vertebra name, inferior and superior side vertebra, verified values including a True/False label for the presence of claw, a True/False label for the presence of traction and a grade of slight, moderate or severe for anterior osteophytes and (2) text files containing 36 coordinates of vertebra boundaries for cervical vertebrae. The verified values for each vertebra and the labeling of the 36 coordinates along the vertebra boundary were provided to NLM by expert radiologists. Based on intersecting the database file providing the verified data and the available text files containing the vertebral boundary coordinates, there were a total of 390 cervical vertebrae for which feature analysis and claw, traction and anterior osteophytes discrimination were performed. For this study, anterior osteophytes discrimination was performed on a normal/abnormal basis, with the inferior and/or superior side of any vertebra labeled as “moderate” or “severe” being called abnormal. Otherwise, the vertebra was labeled as normal for anterior osteophytes discrimination. Finally, in order to evaluate the proposed features, the verified data for the inferior and superior sides for each vertebra was logically Ored. For example, if the inferior side of a vertebra was verified as having claw and the superior side was verified as not having claw, a label of claw was used for the vertebra for classification purposes. Table 2 presents the dataset breakdown of the claw, traction and anterior osteophytes (normal/abnormal) classes for the cervical vertebrae data set that were examined in this study. Included in the cervical vertebrae dataset are 97 C3s, 99 C4s, 96 C5s, 76 C6s and 22 C7s.

Claw, traction, and the presence of anterior osteophytes are based on deviations from the normal square/rectangular shape of a vertebra. The proposed features exploit and quantify such deviations, which can be considered characteristic of claw, trac-

Table 2  
Cervical vertebrae datasets examined for detecting claw, traction and anterior osteophytes

	No. of cervical vertebrae
Claw/no claw	242/148
Traction/no traction	212/178
Anterior osteophytes (abnormal/normal)	258/132 (82 severe, 176 moderate)

tion and anterior osteophytes. The proposed features capture deviations in the region(s) of the vertebra where shape distortion, if it is present, is expected to be detectable. For example, from Fig. 5, the anterior side area captures deviation from normal square/rectangular shape in the region where traction might be observed. This deviation is quantified, relative to the area of the vertebra, in the feature  $F_5$ . The inferior side area provides an indicator of deviation in the region where claw might be observed. This deviation is quantified relative to the area of the vertebra in feature  $F_3$ . Because different vertebrae have different degrees of curvature along the boundary, all of the proposed features described in Section 2.2 were investigated for detecting the three anomalies.

Based on combining the verified data for the inferior and superior sides of the vertebra, as described above, a single label for each of the three anomalies was used for the development of similar but separate classification schemes for the claw, traction and anterior osteophytes cases. Twenty randomly generated training and test sets were generated for claw/no claw, traction/no traction and anterior osteophytes discrimination (abnormal/normal). Ninety percent of the feature vectors for claw and no claw were used for training and the remaining 10% of the feature vectors were utilized as test data. Similar training and test sets were generated for the traction/no traction and abnormal/normal anterior osteophytes cases, where “abnormal osteophytes” refers to moderate and severe cases, as labeled by the expert. Note that the training and test sets do not coincide for each of the three anomalies because the combination of anomalies may differ for different vertebra. The following process was performed for each classification problem (claw/no claw, traction/no traction and abnormal/normal osteophytes). (1) Compute the mean and standard deviation of the features from the training data for each class. (2) Normalize the training data by subtracting the mean and dividing by the standard deviation. (3) Estimate the number of clusters for each class using subtractive clustering [11,12]. (4) Using the training data and estimated number of cluster centers for each class, perform K-means clustering [13] to determine cluster centers for each class. (5) Normalize the test set using the means and standard deviations computed from the training data. (6) Perform nearest neighbor classification for the test data [14]. For each test feature vector, the Euclidean distance is calculated to the cluster centers for each class. The test vector is labeled as belonging to the class for which the test vector has the minimum Euclidean distance. (7) Compute the true positive and true negative classification rates for the test data. The true positive rate is the percentage of correctly classified vertebrae with claw (similar for traction and abnormal vertebrae

for anterior osteophytes), and the true negative rate is the percentage of correctly classified vertebrae with no claw (similar for no traction and normal vertebrae for anterior osteophytes). (8) Repeat steps 1–7 for all 20 randomly generated training and test sets.

Note that the same feature vectors were used for claw/no claw, traction/no traction and abnormal/normal for anterior osteophytes, but the feature vectors were included in the training and test sets based on the feature vector (vertebra class assignments) for claw, traction and anterior osteophytes, respectively. For example, for claw/no claw classification, the makeup of the training and test sets differed from the makeup of the training and test sets for the traction/no traction classification. A slightly modified version of this process was explored for detecting anterior osteophytes. For anterior osteophytes, vertebrae are labeled as having slight, moderate or severe grades. For this study, vertebrae with moderate or severe grades were labeled as abnormal, and vertebrae with slight grades were called normal. In classifier algorithm development, cluster centers were determined for the training data with slight (normal), moderate, and severe grades, separately. Nearest neighbor classification was performed based on finding the minimum Euclidean distance for a given feature vector to the cluster centers determined from the training data for the slight (normal), moderate, and severe grades. The feature vector (vertebra) closest in Euclidean distance to the cluster centers for either the moderate or severe grade cases was classified as abnormal. Otherwise, it was classified as normal.

Four convex hull-based features [8] previously investigated for the detection of anterior osteophytes in lumbar vertebrae were used as a baseline method for comparison. The four convex hull-based features include  $F_1$  and  $F_2$  presented in this paper, the ratio of the sum of the areas of all regions with centroids in the exclusive-OR image X contained on the anterior half of the vertebra to the filled vertebra area  $A_D$ , and the ratio of the largest region with centroid in the exclusive-OR image X contained in the anterior half of the vertebra to the filled vertebra area  $A_D$ . For these feature calculations, the anterior half of the vertebra was determined based on finding the orientation of the vertebra and constructing a dividing line at the orientation angle through the vertebra’s centroid. The anterior side was defined as the region below the dividing line. Fig. 3 shows an image example of the regions within vertebra used for feature calculations.

The feature vectors examined for claw/no claw, traction/no traction and abnormal/normal anterior osteophytes include the six proposed features ( $F_1$ – $F_6$ ), features  $F_1$ – $F_5$  (omitting the “asymmetry index”  $F_6$ ), and the baseline convex hull-based features. Average and standard deviation for the true positive and true negative rates were calculated over the 20 randomly generated training and test sets for analysis and comparison of the features.

#### 4. Results and discussion

This section presents experimental results for claw/no claw, traction/no traction and abnormal/normal osteophyte discrimination for cervical vertebrae data using the experimental approach presented in the previous section.

Table 3

Cervical vertebrae K-means and nearest neighbor test detection results for six convex hull-based features over 20 randomly generated training and test sets

Iteration	% Correct claw	% Correct no claw	% Correct traction	% Correct no traction	% Correct abnormal	% Correct normal
1	84.00	71.43	91.67	80.00	85.19	83.33
2	88.00	78.57	79.17	100.00	92.59	66.67
3	88.00	64.29	83.33	86.67	96.30	66.67
4	80.00	78.57	87.50	93.33	77.78	83.33
5	84.00	71.43	83.33	93.33	81.48	83.33
6	80.00	78.57	100.00	73.33	92.59	66.67
7	84.00	92.86	79.17	100.00	85.19	83.33
8	96.00	57.14	79.17	100.00	81.48	66.67
9	88.00	71.43	83.33	86.67	85.19	66.67
10	88.00	78.57	87.50	80.00	85.19	66.67
11	92.00	64.29	91.67	73.33	81.48	83.33
12	80.00	78.57	83.33	86.67	92.59	83.33
13	84.00	71.43	83.33	86.67	77.78	75.00
14	80.00	78.57	83.33	86.67	88.89	83.33
15	84.00	85.71	83.33	86.67	77.78	83.33
16	96.00	57.14	83.33	86.67	85.19	83.33
17	76.00	85.71	91.67	66.67	81.48	75.00
18	92.00	71.43	87.50	73.33	74.07	91.67
19	84.00	85.71	83.33	80.00	81.48	83.33
20	88.00	64.29	95.83	60.00	85.19	66.67
Mean	85.80	74.29	86.04	84.00	84.44	77.08
Standard deviation	5.43	9.67	5.62	10.90	5.84	8.50

Table 3 presents the cervical vertebrae test results for the claw/no claw, traction/no traction and abnormal/normal for anterior osteophytes cases using the clustering technique presented above with the six proposed features over 20 randomly generated training and test sets. These experiments were performed using 390 cervical vertebrae with the distribution of claw/no claw, traction/no traction and abnormal (moderate and severe)/normal for anterior osteophytes, as described in Table 2. In Table 3, the first column shows the randomly generated training and test set iteration number. Columns 2 and 3 present the claw/no claw results. Columns 4 and 5 give the traction/no traction results. Columns 5 and 6 show the abnormal/normal results for detecting anterior osteophytes. Vertebrae labeled moderate or severe are considered abnormal. Vertebrae labeled as slight are taken as normal.

Table 4 gives the number of clusters used determined using the subtractive clustering technique from the training data for K-means clustering and nearest neighbor classification over the 20 training and test sets. Note that the vertebrae used in the training and test sets are different for the claw/no claw, traction/no traction and abnormal/normal anterior osteophytes cases because each vertebra has different claw, traction or osteophytes characteristics.

Table 5 presents the average cervical vertebrae test results for the claw/no claw, traction/no traction and abnormal/normal for anterior osteophytes cases using the clustering technique presented above with five features ( $F_1$ – $F_5$ ), leaving out the moment normalization feature  $F_6$ , over 20 randomly generated training and test sets. These experiments were performed using 390 cervical vertebrae with the distribution of claw/no claw, traction/no traction and abnormal (moderate and severe)/normal for anterior osteophytes, as shown in Table 2. Columns 2 and 3 present the claw/no claw results. Columns 4 and 5 give the traction/no

traction results. Columns 5 and 6 show the abnormal/normal results for detecting anterior osteophytes. As mentioned earlier, vertebrae labeled moderate or severe are considered abnormal, and those labeled as slight are taken as normal.

Table 6 presents the average cervical vertebrae claw/no claw, traction/no traction and abnormal/normal anterior osteophytes test results over the same 20 randomly generated training and test sets using the baseline convex hull-based features [8] as a benchmark for comparison to the extensions developed for the convex hull-based features.

The results presented in Tables 3–6 provide for several observations. First, from Tables 3–5, the six-feature case ( $F_1$ – $F_6$ ) provided higher average correct results than the five-feature case ( $F_1$ – $F_5$ ) and the baseline convex hull-based features. Table 7 summarizes the average percentage improvement of the six size-invariant features over the five feature and baseline convex hull-based feature cases for claw/no claw, traction/no traction and abnormal/normal osteophytes. From Table 7, the proposed features ( $F_1$ – $F_6$ ) consistently outperformed the baseline convex hull-based features for discrimination of claw, traction, and anterior osteophytes. The baseline convex hull-based features were developed to highlight protrusion regions on the anterior side of the vertebra, characteristic of osteophytes. The proposed features provide an alternative approach to highlight the protrusion regions, emphasizing the deviation of a vertebra’s shape from the typical rectangular shape.  $F_1$  and  $F_2$  highlight the deviation of a vertebra’s shape from the rectangular shape based on the size of concave regions around the vertebra.  $F_3$ – $F_5$  provide measures of the concavity of the vertebra on the inferior, superior and anterior sides, respectively, which together can be used to localize protrusion regions characteristic of claw, traction and/or anterior osteophytes. For instance, high ratios for  $F_3$  and  $F_5$  for a vertebra may indicate that there is a protrusion region on the inferior side

Table 4  
Number of cluster centers used for K-means clustering for nearest neighbor classification for claw/no claw, traction/no traction and abnormal/normal anterior osteophytes

Iter.	No. of clusters claw	No. of clusters no claw	No. of clusters traction	No. of clusters no traction	No. of clusters severe osteophytes	No. of clusters moderate osteophytes	No. of clusters normal
1	4	4	3	6	8	4	4
2	4	5	3	5	8	3	3
3	3	5	3	6	6	4	3
4	4	4	3	5	5	4	3
5	4	6	3	6	8	4	5
6	4	4	3	8	7	4	3
7	4	6	3	5	8	5	4
8	3	4	3	5	6	3	4
9	4	4	4	5	9	3	4
10	4	5	3	8	6	3	5
11	3	5	3	5	8	3	4
12	4	4	3	5	6	3	4
13	4	6	4	6	6	6	4
14	4	4	3	5	6	4	4
15	4	6	3	5	7	3	4
16	3	4	3	5	7	4	3
17	4	5	2	5	6	3	3
18	4	6	2	5	8	4	4
19	4	5	3	7	7	4	3
20	4	5	3	8	7	3	3

Note that abnormal osteophytes is based on combining the severe and moderate osteophytes cases.

Table 5  
Cervical vertebrae K-means and nearest neighbor test detection results for five convex hull-based features over 20 randomly generated training and test sets

	Claw	No claw	Traction	No traction	Abnormal osteophytes	Normal
Mean % correct	85.20	70.70	81.30	78.00	86.30	65.80
Standard deviation	7.90	13.31	7.09	10.84	6.37	10.44

Table 6  
Cervical vertebrae K-means and nearest neighbor test detection results for baseline convex hull-based features [8] over 20 randomly generated training and test sets

	Claw	No claw	Traction	No traction	Abnormal osteophytes	Normal
Mean % correct	63.00	78.20	77.90	62.10	81.70	60.52
Standard deviation	8.70	15.90	10.91	15.16	8.10	11.62

of the vertebra. A high value for  $F_5$  with relatively low values for  $F_3$  and  $F_4$ , respectively, may be an indicator of a protrusion region on the anterior face of the vertebra, which could be characteristic of traction. Part of the rationale for introducing feature  $F_6$  was to quantify the relative symmetry between the anterior and posterior sides of the vertebra in order to identify potential protrusion regions which may be convex. This feature appears to contribute to successful discrimination of claw, traction and osteophytes, providing 2.10%, 5.37% and 4.71% improvement

of claw, traction and osteophytes discrimination when combined with features  $F_1-F_5$ .

Second, from Table 3, the proposed features  $F_1-F_6$  yielded the highest average correct results for discriminating traction/no traction, with an overall average of 85.02% (average of traction and no traction results). This compares to an overall average of 80.05% for claw/no claw and 80.76% for abnormal/normal osteophytes. The features  $F_1-F_6$  are able to highlight protrusion regions and asymmetrical shape that is often characteristic

Table 7  
Percentage improvement in average claw/no claw, traction/no traction and abnormal/normal anterior osteophytes between  $F_1-F_6$  features and  $F_1-F_5$  and baseline convex hull-based features

	% Improvement claw/no claw	% Improvement traction/no traction	% Improvement abnormal/normal osteophytes
Features $F_1-F_5$	2.10	5.37	4.71
Baseline convex hull-based features	9.45	15.02	9.66

Table 8

$p$ -Value associated with the six convex hull-based features for the claw, traction and anterior osteophyte cases

Serial number	Name of feature	$p$ -Value		
		Claw	Traction	Anterior osteophyte
1	$F_1$	<b>0.0048</b>	0.7699	0.4207
2	$F_2$	<b>0.0002</b>	0.1946	0.5127
3	$F_3$	<b>0.0004</b>	<b>0.0025</b>	0.2024
4	$F_4$	0.8064	<b>0.0252</b>	<b>0.0723</b>
5	$F_5$	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>0.0039</b>
6	$F_6$	0.3103	<b>0.0001</b>	<b>0.0033</b>

The statistically significant features for each pathological condition have been printed in bold.

of claw, traction and osteophytes.  $F_6$  provides the capability to quantify the location of protrusion regions relative to the vertebra's shape and inherent curvature, as manifested on the posterior side of the vertebra.

Logistic regression [15] was performed on the six features ( $F_1$ – $F_6$ ) for the cervical spine data set to determine which features are statistically significant in the discrimination of claw, traction and anterior osteophytes, respectively. Table 8 shows the  $p$ -value from the regression analysis associated with each of the six proposed features ( $F_1$ – $F_6$ ) for discriminating claw, traction and anterior osteophytes in cervical vertebrae. A  $p$ -value of less than or equal to 0.05 implies that the associated feature is statistically significant in the discrimination of a specific abnormality [15]. From Table 8, it is observed that the features  $F_1$ – $F_3$  and  $F_5$  are statistically significant for detecting the presence of claw in cervical vertebrae. The features  $F_3$ – $F_6$  are statistically significant for detecting traction, and anterior osteophytes can be best identified by the features  $F_4$ – $F_6$ . Therefore, it can be inferred that the feature  $F_6$  is of greater significance to the detection of traction and anterior osteophytes. This is in agreement with the results shown in Table 7 where, the addition of the feature  $F_6$  to the reduced feature set  $F_1$ – $F_5$  resulted in the least improvement (2.10%) in the average correct classification for the claw/no claw case. Feature  $F_6$  involved using moment normalization for computing the vertebra's area and exclusive-OR area of the flipped anterior side about the axis through the centroid of the vertebra. Moment normalization was used in order to mitigate the effect of vertebra orientation and skewness in shape.

Finally, the experimental results show the potential for the proposed size-invariant features to identify key vertebra attributes such as claw, traction and osteophytes. The same six features have achieved comparable results in detecting the three vertebrae attributes, with slightly better recognition results for traction. The features  $F_1$ – $F_6$  were computed based on a vertebra representation of 36 points. In previous research, the baseline convex hull-based features were developed based on 55-point vertebra representations [8], providing slightly more detail in representing the vertebra boundary. For this research, the K-means clustering and nearest neighbor approach labeled vertebrae as containing claw, traction or osteophytes based on separately trained models. Classification was performed using

expert grading of vertebrae for the presence of claw, traction, and osteophytes.

## 5. Conclusions

The overall goal of this research is the development of features and techniques that can be incorporated into a content-based image retrieval system to facilitate the query of images containing vertebrae with specific types of conditions related to osteoarthritis. New size-invariant descriptors were investigated for the discrimination of claw, traction, and anterior osteophytes in cervical vertebrae. Using a K-means and nearest neighbor technique, test results yielded correct average discrimination rates of 85.80%, 86.04% and 84.44% for claw, traction, and anterior osteophytes, respectively, with corresponding average correct normal vertebrae discrimination rates of 74.29%, 84.00% and 77.08%, respectively. Logistic regression analysis of the proposed features shows that the proposed features contain statistically significant information for the discrimination of claw, traction and anterior osteophytes. Exploring fuzzy logic-based techniques to provide different degrees of association or membership with each of the attributes is currently under investigation to improve classification capability.

## Acknowledgments

This work was supported by National Library of Medicine under contract number 467-MZ-601342 and the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

## References

- [1] Xu X, Lee DJ, Antani SK, Long LR. Partial shape matching of spine X-ray shapes using dynamic programming. In: Proceedings of the 17th IEEE symposium on computer-based medical systems, vol. 17. 2004. p. 97-02.
- [2] Lee DJ, Antani SK, Xu X, Long LR. Design and evaluation of a curve matching based X-ray image retrieval system. In: Proceedings of the SPIE medical imaging 2005: PACS and imaging informatics, vol. 5748. 2005. p. 365-73.
- [3] Heggeness MH, Doherty BJ. Morphologic study of lumbar vertebral osteophytes. South Med J 1998;91(2):187-9.
- [4] Pate D, Goobar J, Resnick D, Haghighi P, Sartoris DJ, Pathria MN. Traction osteophytes of the lumbar spine: radiographic-pathologic correlation. Radiology 1988;166(3):843-6.
- [5] Xu X, Antani SK, Lee DJ, Long LR, Thoma GR. Relevance feedback for shape-based pathology in spine X-ray image retrieval. In: Proceedings of the SPIE medical imaging 2006: PACS and imaging informatics. 2006. p. 6145.
- [6] Antani SK, Long LR, Thoma GR. Applying vertebral boundary semantics to CBIR of digitized spine X-ray images. Proceedings of the SPIE storage and retrieval methods and applications for multimedia, vol 5682 2005:98-107.
- [7] De Bruijine M, Nielsen M. Lecture notes in computer science medical image computing and computer-assisted intervention. In: Proceedings of the medical image computing and computer-assisted intervention: MICCAI 2004: 7th international conference, vol. 3216, no. 1. 2004. p. 168-75.
- [8] Cherkuri M, Stanley RJ, Long LR, Antani SK, Thoma GR. Anterior osteophyte discrimination in lumbar vertebrae using size-invariant features. Comput Med Imaging Graph 2004;28(1/2):99-108.

- [9] Barber CB, Dobkin DP, Huhdanpaa HT. The Quickhull algorithm for convex hulls. *ACM Trans Math Softw* 1996;22(4):469–83.
- [10] Horn BK. *Robot vision*. 1st ed. MIT Press; 1986.
- [11] Chiu S. Fuzzy model identification based on cluster estimation. *Intell Fuzzy Syst J* 1994;2(3):267–78.
- [12] Yager R, Filev D. Generation of fuzzy rules by mountain clustering. *Intell Fuzzy Syst J* 1994;2(3):209–19.
- [13] Seber GAF. *Multivariate observations*. Wiley; 1984.
- [14] Shakhnarovich G, Darrell T, Indyk P. *Nearest-neighbor methods in learning and vision*. 1st ed. MIT Press; 2006.
- [15] Cary NC. *Logistic regression examples using the SAS system*. SAS Institute Inc.; 1995.

**Ronald Joe Stanley** is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Missouri-Rolla. His research interests include signal and image processing, pattern recognition and automation. He is a Senior Member of the IEEE and a member of NAFIPS. He received the B.S.E.E. and M.S.E.E. degrees in electrical engineering and a Ph.D. degree in Computer Engineering and Computer Science from the University of Missouri-Columbia. As a graduate student at the University of Missouri-Columbia, he worked under training grants from the National Library of Medicine and the National Cancer Institute. Upon completing his doctoral study, he served as Principal Investigator for the Image Recognition program at Systems & Electronics Inc. in St. Louis, MO.

**Sameer Antani** is a Staff Scientist with the Lister Hill National Center for Biomedical Communications at the National Library of Medicine a part of the U.S. National Institutes of Health. He conducts research on various topics in content-based image retrieval, medical multimedia databases, next-generation interactive documents, and advanced multimodal medical document retrieval. He earned his Master of Engineering and Ph.D. in Computer Science and Engineering from the Pennsylvania State University in 1998 and 2001, respectively. He earned his Bachelors degree in Computer Engineering from the University

of Pune, India, in 1994. Dr. Antani is a member of the IEEE and the IEEE Computer Society.

**L. Rodney Long** is an electronics engineer for the Communications Engineering Branch in the Lister Hill National Center for Biomedical Communications, a research arm of the National Library of Medicine. He has held his current position since 1990. Prior to his current job, he worked for 14 years in industry as a software developer and as a systems engineer. His research interests are in image processing and scientific/biomedical databases, with emphasis on the content-based retrieval of medical images. He has an M.A. in applied mathematics from the University of Maryland.

**George R. Thoma** is Chief of the Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications, a research and development division of the U.S. National Library of Medicine. In this capacity, he directs R&D programs in document image analysis and understanding, biomedical image processing, image compression, automated document image delivery, digital X-ray archiving, animated virtual books, and high speed image transmission. He earned a B.S. from Swarthmore College, and the M.S. and Ph.D. from the University of Pennsylvania, all in electrical engineering. Dr. Thoma is a Fellow of the SPIE, the International Society for Optical Engineering.

**Kapil Gupta** is currently pursuing a Ph.D. in Electrical Engineering at the University of Missouri-Rolla. He received his primary and secondary education in Calcutta, India. He received his Bachelor's degree in Electronics and Telecommunications Engineering in July 2001 from Amravati University, India. He received his Master's degree in Electrical Engineering from the University of Missouri-Rolla in December 2003.

**Mohammed Das** is currently pursuing an M.S. degree in Computer Science at the University of Missouri-Rolla. He received his primary and secondary education in Bombay, India. He received his Bachelor of Engineering in Information Technology in June 2006 from the University of Mumbai, India.